



# Faculty of Computers and Artificial Intelligence

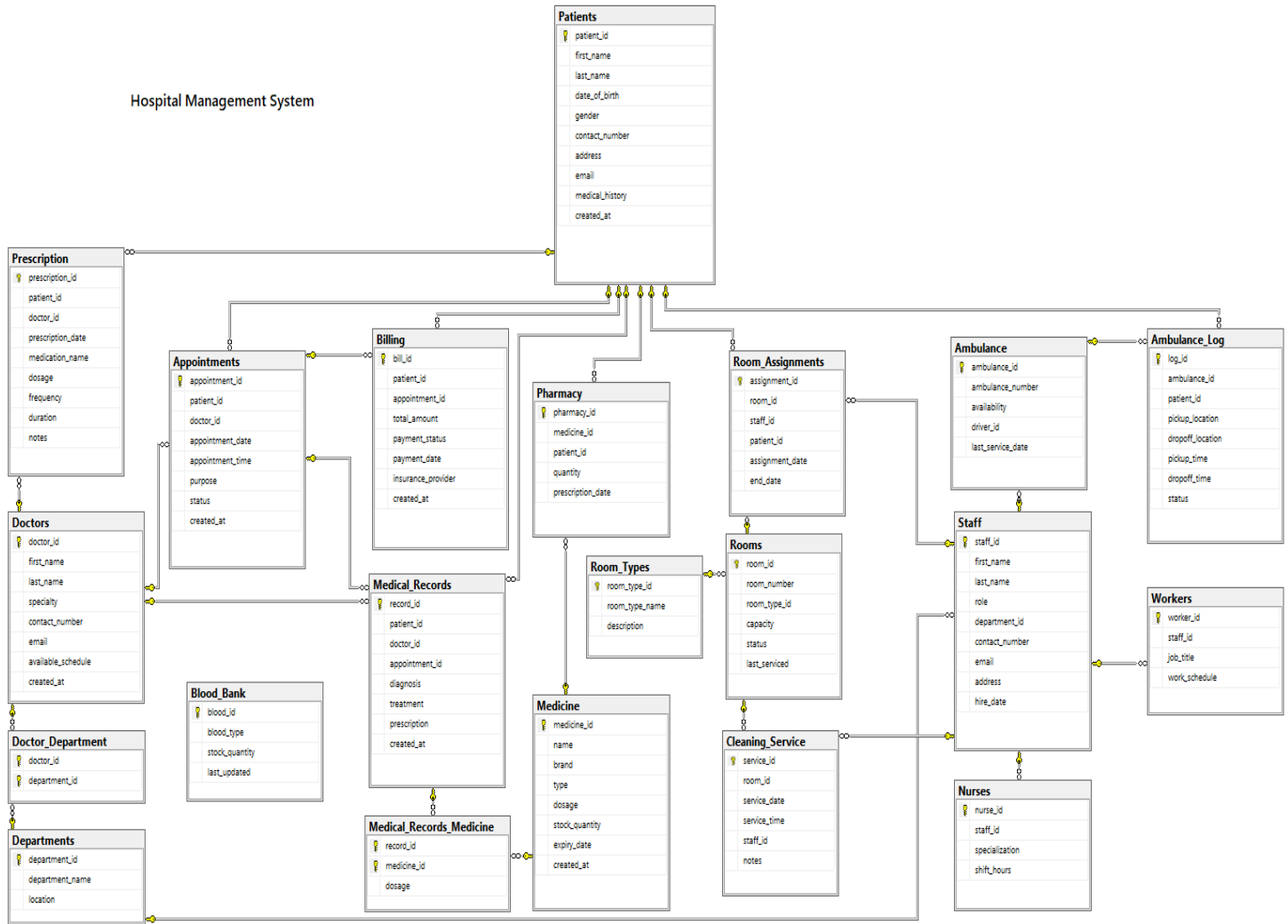
## Data Warehousing – May 2025

### Hospital Data Warehouse Project

#### 1. Team Members

	Name	ID	Group Number
1-	Abdelrahman Mostafa Sayed	20220197	3IS-S7
2-	Mariam Assem	20221142	3IS-S2
3-	Hager Omar	20221190	3IS-All
4-	Reem Ahmed Abdallah	20221063	3IS-S2
5-	Samar Mahmoud	20221076	3IS-S2

## 2. Physical model of the source system:



### **3a. Define the business processes that you will model:**

#### **1. Fact\_Pharmacy – Medication Dispensing Process:**

- Tracks the distribution and usage of prescribed medicine.
- KPIs:
  - Total quantity dispensed by medicine type and year
  - Most prescribed medicines per year
  - Unique patient count per medicine

#### **2. Fact\_Appointments – Patient Visit Process**

- Tracks patient visits and interactions with doctors.
- KPIs:
  - Top No of Appointments count per patient per year
  - Average appointments per doctor per year
  - Top Doctors by Appointment Volume

#### **3. Fact\_Billing – Revenue Collection Process**

- Captures all billing transactions and financial flow.
- KPIs:
  - Average billing per patient per year
  - Total revenue by year
  - Monthly revenue trends
  - Top Patients by Total Billing

#### **4. Fact\_Cleaning\_Service – Facility Hygiene Management Process**

- Tracks room cleaning activities and workload per staff.
  - KPIs:
    - Total No. of cleaning services done by each staff member in each year
    - Yearly Cleaning Services per Room
    - Total services by room type
-

### **3b. Declare the grain of each fact table:**

The grain of each fact table is defined at the most detailed (minimum) level of the respective business process. This minimum grain allows flexible aggregation and supports detailed analysis across multiple dimensions.

<b>Fact Table</b>	<b>Grain (One row represents...)</b>
<b>Fact_Pharmacy</b>	One medication dispensed to a patient on a specific date.
<b>Fact_Appointments</b>	One appointment made by one patient with one doctor on a specific date .
<b>Fact_Billing</b>	One bill issued to one patient on a specific date
<b>Fact_Cleaning_Service</b>	Number of cleaning services performed by one staff member in one room on a specific date

### **3c. Define the type of each fact table:**

<b>Fact Table</b>	<b>Fact Table Type</b>	<b>Reasoning</b>
<b>Fact_Pharmacy</b>	Transaction Fact Table	Each row represents a medication dispense event.
<b>Fact_Appointments</b>	Factless Fact Table	Represents appointment events with no measures. Counting only.( Each row = a distinct doctor-patient appointment)
<b>Fact_Billing</b>	Transaction Fact Table	Each record is a single billing transaction for a specific patient.
<b>Fact_Cleaning_Service</b>	Periodic Snapshot	Each row = a summary of cleanings, not individual cleanings (count of cleaning)

### 3d. Define the dimensions and the type of each one:

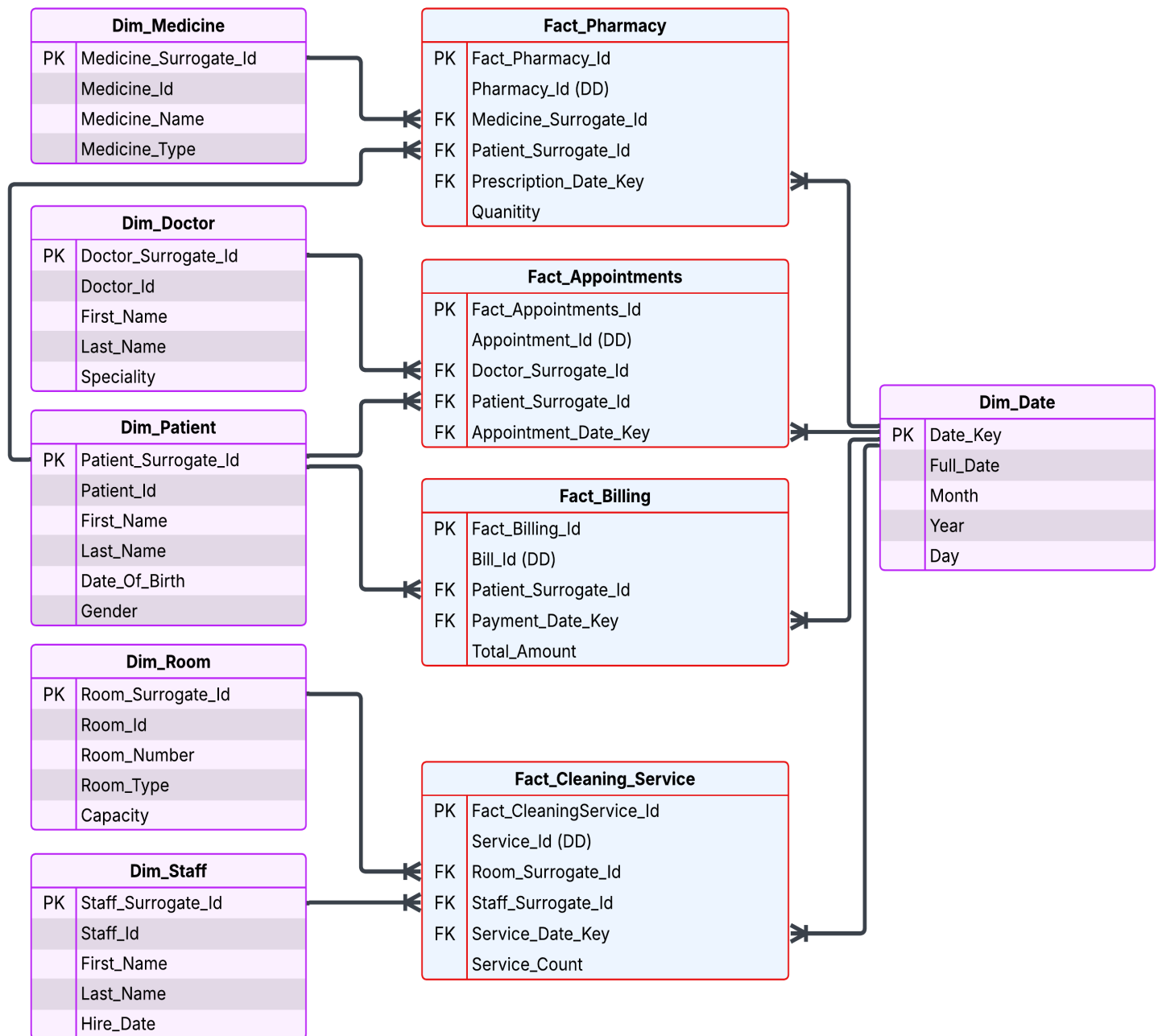
<b>Dimension Table</b>	<b>Type</b>	<b>Reasoning</b>
<b>Dim_Patient</b>	<b>Conformed</b>	Used across multiple fact tables like <b>Fact_Appointments</b> , <b>Fact_Billing</b> , <b>Fact_Pharmacy</b> .
<b>Dim_Date</b>	<b>Conformed, Role-Playing .</b>	Used across all facts for date-related analysis , and reused with different date roles (appointment date, billing date, cleaning service date)
<b>Dim_Room</b>	<b>Static Dimension</b>	Room number, type, and capacity don't typically change often
<b>Dim_Staff</b>	<b>Slowly Changing Dimension (SCD)</b>	staff attributes like <b>First_Name</b> , <b>Last_Name</b> may change over time
<b>Dim_Medicine</b>	<b>Static Dimension</b>	Medicine name and type are usually fixed; changes are rare and often treated by introducing a new medicineID.
<b>Dim_Doctor</b>	<b>Slowly Changing Dimension (SCD)</b>	'Speciality', 'First_Name', 'Last_Name ' attribute may be change over time
<b>Dim_Pharmacy</b>	<b>Degenerate Dimension</b>	It has ONLY primary key attribute "Pharmacy_Id" and this attribute is stored in the fact table "Fact_Pharmacy" and has no associated dimension table

<b>Dim_Appointment</b>	<b>Degenerate Dimension</b>	It has <b>ONLY</b> primary key attribute “Appointment_Id” and this attribute is stored in the fact table “Fact_Appointments” and has no associated dimension table
<b>Dim_Bill</b>	<b>Degenerate Dimension</b>	It has <b>ONLY</b> primary key attribute “Bill_Id” and this attribute is stored in the fact table “Fact_Billing” and has no associated dimension table
<b>Dim_Service</b>	<b>Degenerate Dimension</b>	It has <b>ONLY</b> primary key attribute “Service_Id” and this attribute is stored in the fact table “Fact_Cleaning_Service” and has no associated dimension table

**3e. Define the measures that will appear in the fact tables and the type of each one:**

<b>Fact Table</b>	<b>Measure</b>	<b>Measure Type</b>
<b>Fact_Pharmacy</b>	<b>Quantity</b>	<b>Fully Additive</b> - Can be summed across all dimensions including time.
<b>Fact_Billing</b>	<b>Total_Amount</b>	<b>Fully Additive</b> - Can be summed over time, patients.
<b>Fact_Cleaning_Service</b>	<b>Service_Count</b>	<b>Fully Additive</b> - Number of cleanings can be summed across all dimensions.

### 3f. Physical model of DWH (the final star/galaxy schema):



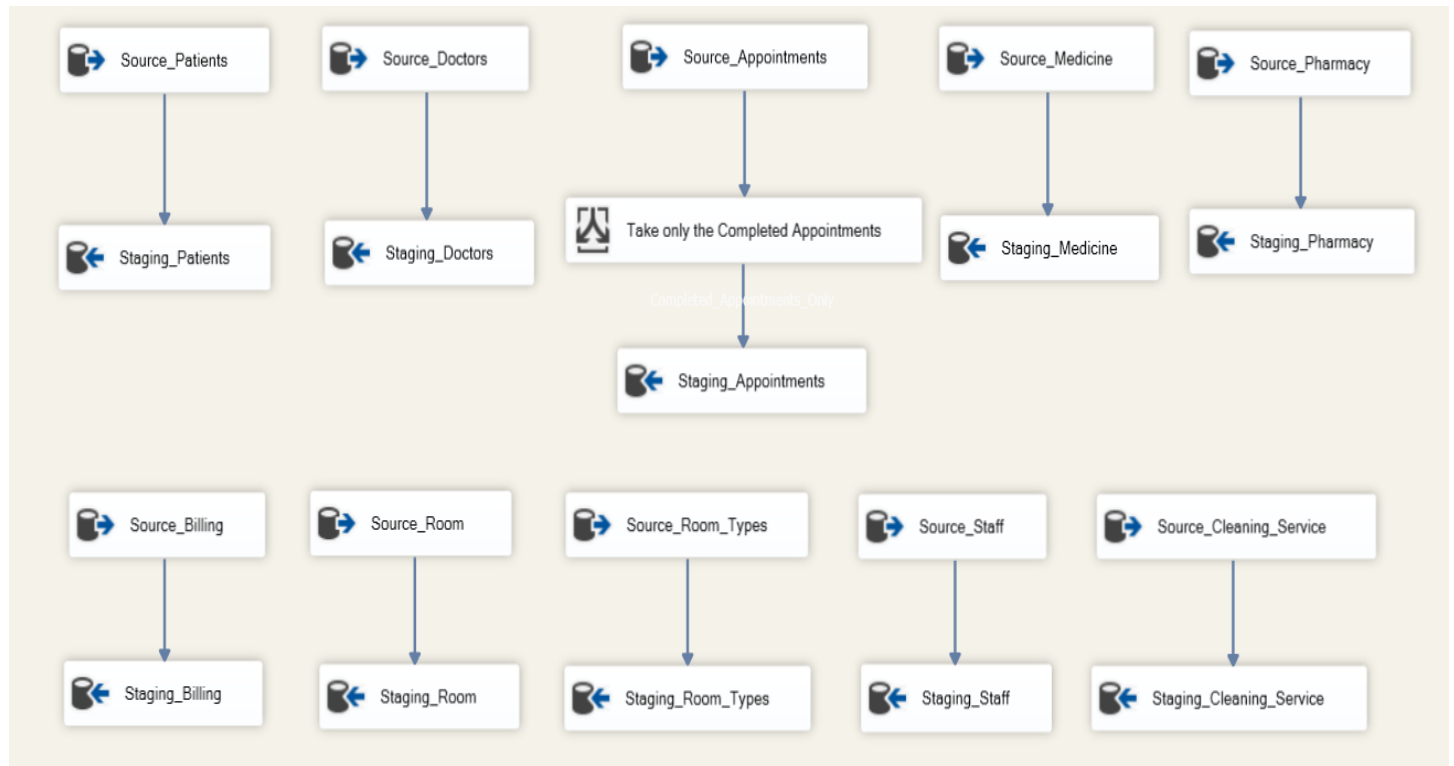
#### 4. Screenshots of the data flow tasks, and control flow tasks used for building the DWH:

##### **Control Flow for the Whole Package:**

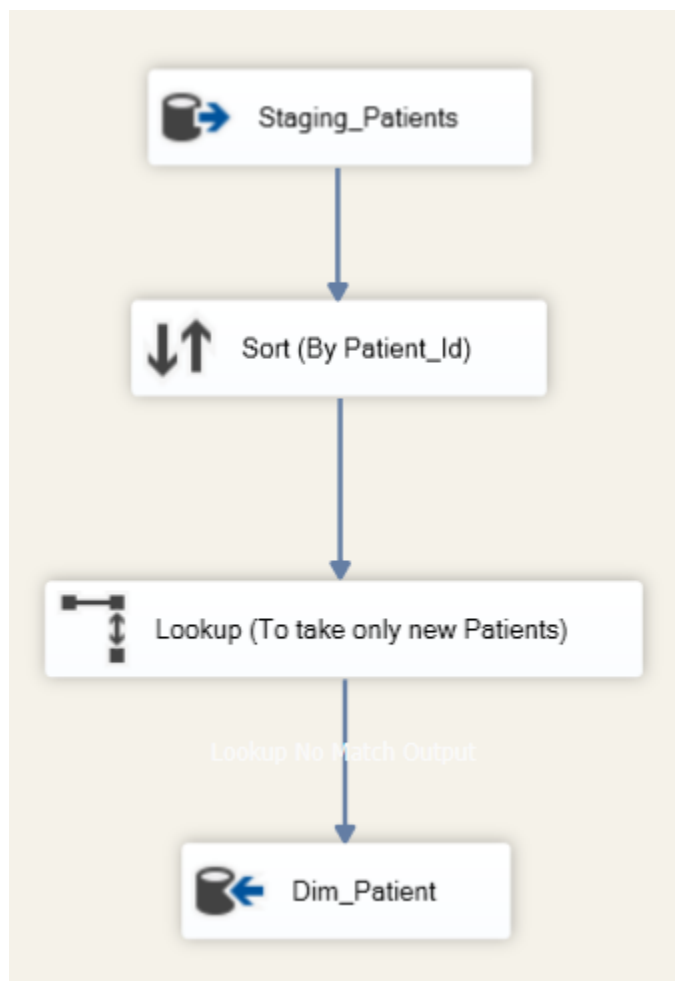




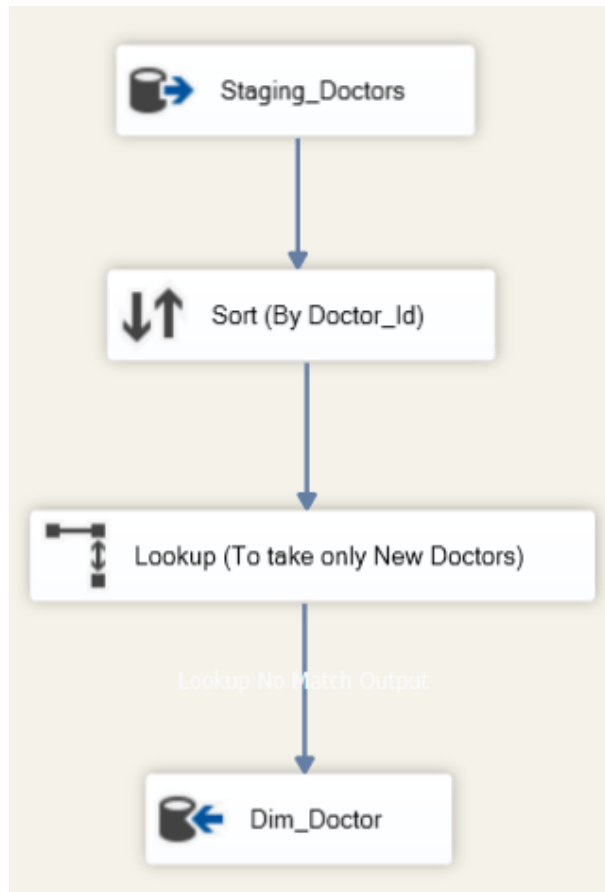
## Data Flow Task for Load Staging Tables:



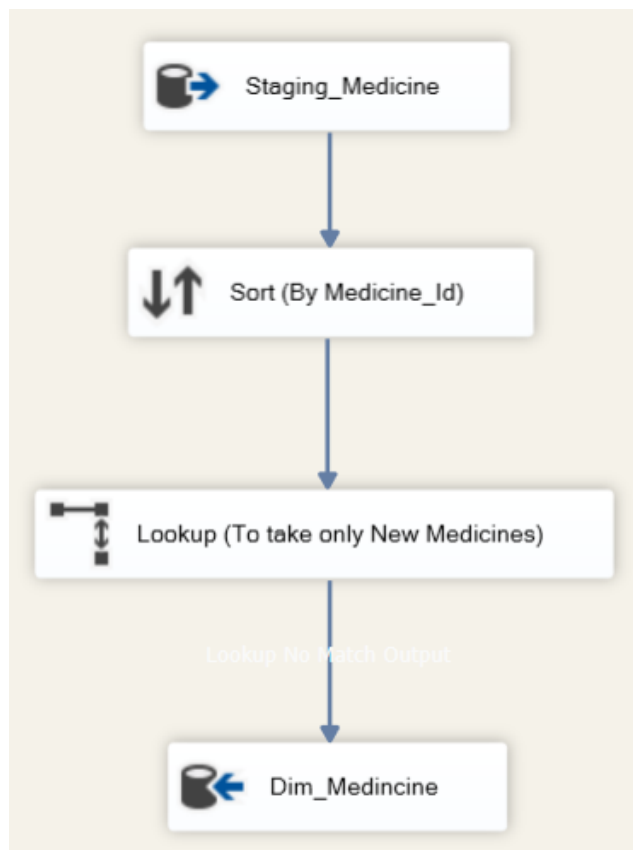
## Data Flow Task for Load Dim\_Patients:



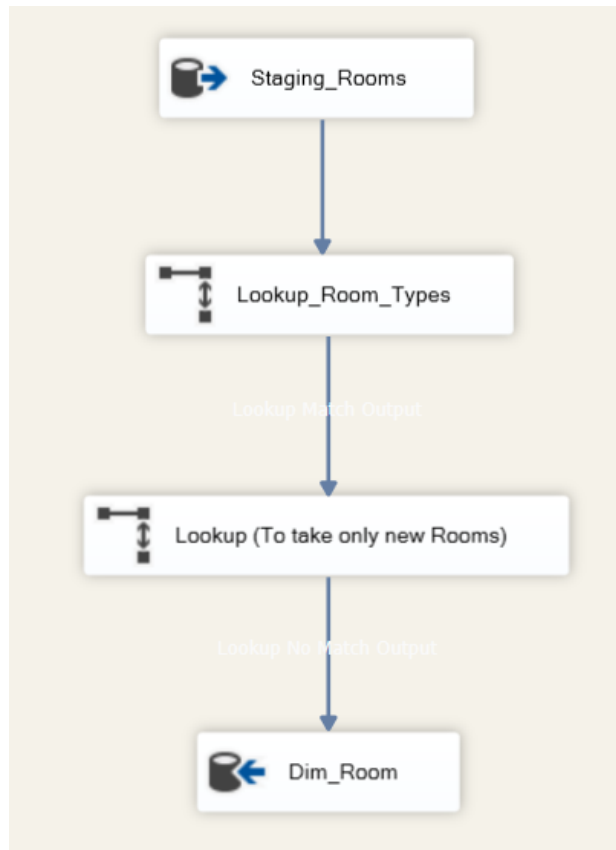
## Data Flow Task for Load Dim\_Doctor:



## Data Flow Task for Load Dim\_Medicine:



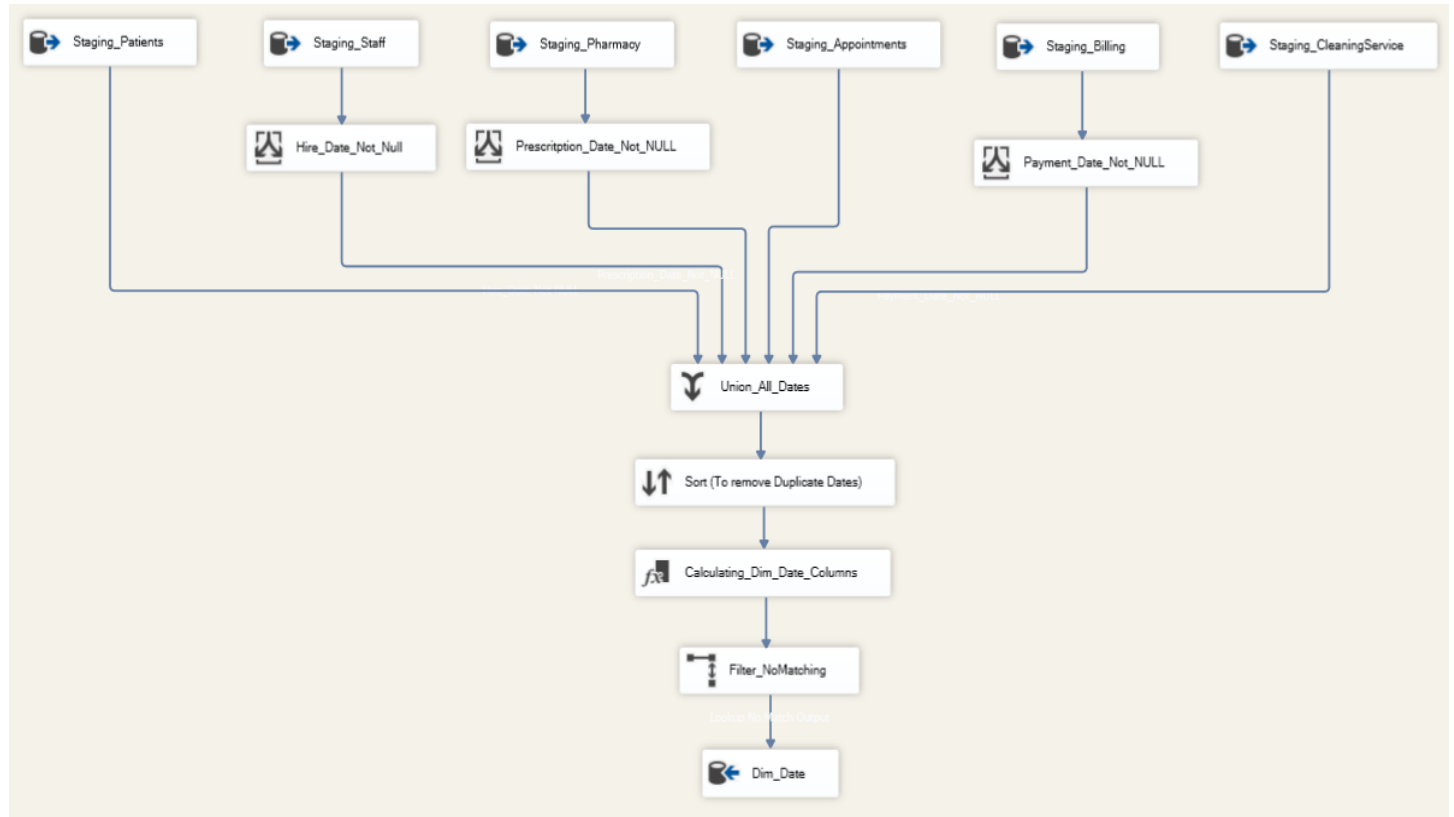
## Data Flow Task of Load Dim\_Room:



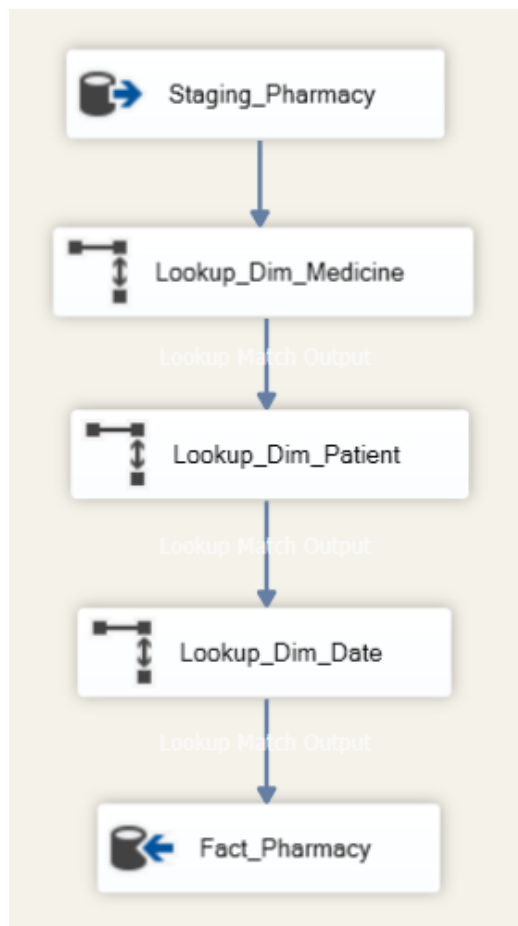
## Data Flow Task for Load Dim\_Staff:



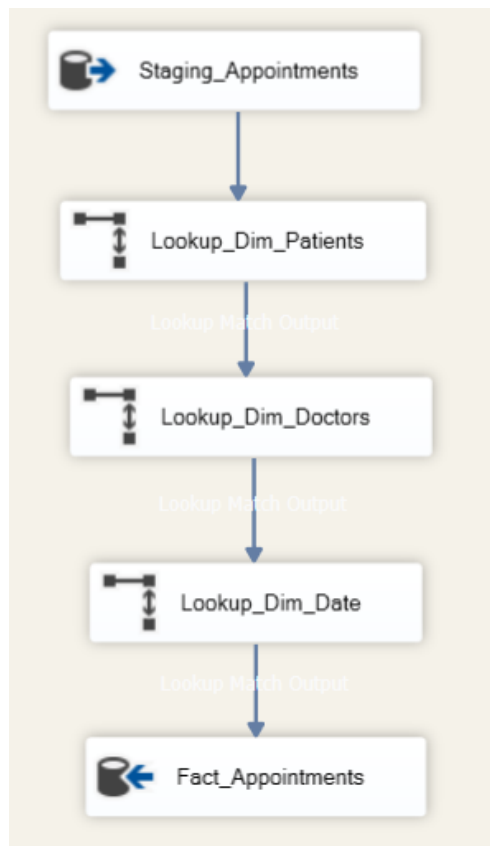
## Data Flow Task for Load Dim\_Date:



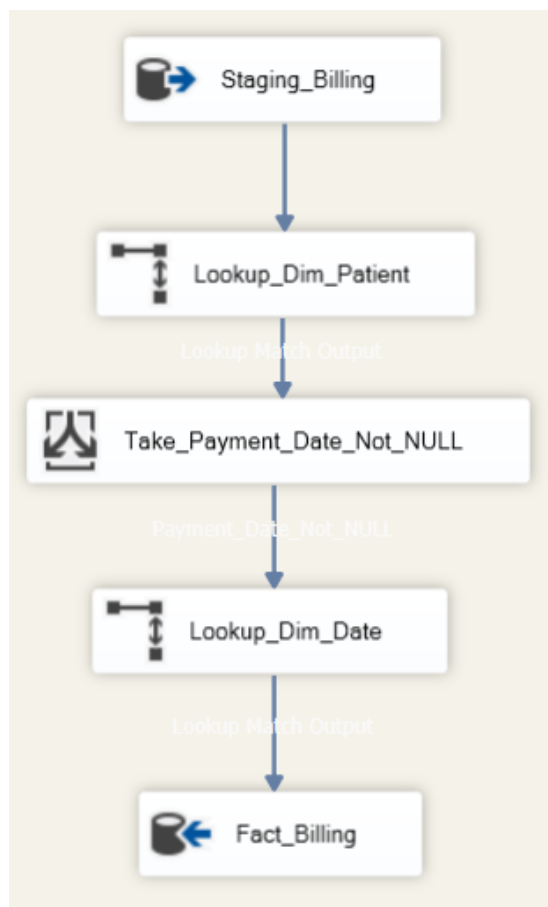
## Data Flow Task for Load Fact\_Pharmacy:



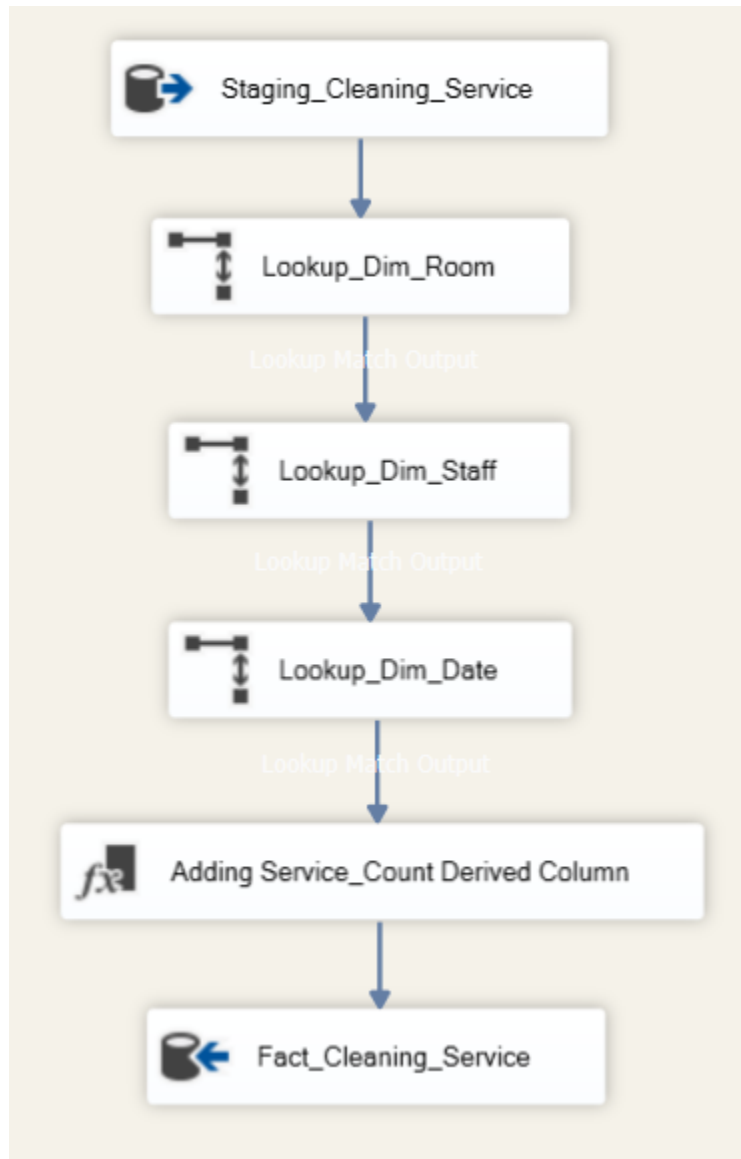
## Data Flow Task for Load Fact\_Appointments:



## Data Flow Task for Load Fact\_Billing:



## Data Flow Task for Load Fact\_Cleaning\_Service:



## 5. Queries on each fact table to let me understand what this fact table represents and what insights we can get from it, and a screenshot of the result set for each query:

### Fact\_Pharmacy:

```
-- 1. Quantity Dispensed by Medicine Type
SELECT
    dm.Medicine_Type,
    SUM(fp.Quantity) AS Total_Quantity_Dispensed
FROM
    Fact_Pharmacy fp
    INNER JOIN Dim_Medicine dm ON fp.Medicine_Surrogate_Id = dm.Medicine_Surrogate_Id
GROUP BY
    dm.Medicine_Type
ORDER BY
    Total_Quantity_Dispensed DESC;
```

	Medicine_Type	Total_Quantity_Dispensed
1	Tablet	7985
2	Capsule	6285
3	Ointment	630
4	Injection	435
5	Liquid	145

```
-- 2. Top Prescribed Medicines per Year
SELECT
    dd.Year,
    dm.Medicine_Name,
    SUM(fp.Quantity) AS Total_Quantity
FROM
    Fact_Pharmacy fp
    INNER JOIN Dim_Medicine dm ON fp.Medicine_Surrogate_Id = dm.Medicine_Surrogate_Id
    INNER JOIN Dim_Date dd ON fp.Prescription_Date_Key = dd.Date_Key
GROUP BY
    dd.Year,
    dm.Medicine_Name
ORDER BY
    dd.Year ASC,
    Total_Quantity DESC;
```

	Year	Medicine_Name	Total_Quantity
1	2022	Donepezil	270
2	2022	Venlafaxine	150
3	2022	Metoprolol	120
4	2022	Esomeprazole	120
5	2022	Levofloxacin	120
6	2022	Amlodipine	110
7	2022	Methylprednisolone	105
8	2022	Atorvastatin	100
9	2022	Metformin	100
10	2022	Furosemide	100
11	2022	Ranitidine	95
12	2022	Methotrexate	90

Query executed successfully.

```
-- 3. Number of Patients per Medicine
```

```
SELECT
    dm.Medicine_Name,
    COUNT(DISTINCT fp.Patient_Surrogate_Id) AS Patient_Count
FROM
    Fact_Pharmacy fp
    INNER JOIN Dim_Medicine dm ON fp.Medicine_Surrogate_Id = dm.Medicine_Surrogate_Id
GROUP BY
    dm.Medicine_Name
ORDER BY
    Patient_Count DESC;
```

	Medicine_Name	Patient_Count
1	Levothyroxine	26
2	Venlafaxine	22
3	Tamsulosin	18
4	Paracetamol	14
5	Doxorubicin	14
6	Amoxicillin	12
7	Lidocaine	12
8	Donepezil	6
9	Atorvastatin	6
10	Citalopram	6
11	Doxycycline	5
12	Isoniazid	5

Query executed successfully. MATRIX (16.0 RTM) MATRIX\Abdelrahman Mat... Hosp

## Fact\_Appointments:

```
-- 1. Number of Appointments per each Patient in each year
```

```
SELECT
    dd.Year,
    dp.Patient_Surrogate_Id,
    CONCAT(dp.first_name, ' ', dp.last_name) AS Patient_Full_Name,
    COUNT(*) AS Appointments_Count
FROM
    Fact_Appointments fa
    INNER JOIN Dim_Patient dp ON fa.Patient_Surrogate_Id = dp.Patient_Surrogate_Id
    INNER JOIN Dim_Date dd ON fa.Appointment_Date_Key = dd.Date_Key
GROUP BY
    dd.Year,
    dp.Patient_Surrogate_Id,
    CONCAT(dp.first_name, ' ', dp.last_name);
```

	Year	Patient_Surrogate_Id	Patient_Full_Name	Appointments_Count
1	2022	1	John Doe	1
2	2022	2	Jane Smith	1
3	2022	3	Michael Johnson	1
4	2022	4	Emily Davis	1
5	2022	5	Daniel Brown	1
6	2022	6	Olivia Martinez	1
7	2022	7	James Wilson	1
8	2022	8	Sophia Moore	1
9	2022	9	David Taylor	1
10	2022	10	Isabella Anderson	1
11	2022	11	Fatima Hussein	1
12	2022	12	Juan Rodriquez	1

Query executed successfully.



-- 2. Average Number of Appointments per Doctor in each year

```
SELECT
    ddate.Year,
    fa.Doctor_Surrogate_Id,
    CONCAT(dd.first_name, ' ', dd.last_name) AS Doctor_Full_Name,
    COUNT(*) AS Appointments_Count
FROM
    Fact_Appointments fa
    INNER JOIN Dim_Doctor dd ON fa.Doctor_Surrogate_Id = dd.Doctor_Surrogate_Id
    INNER JOIN Dim_Date ddate ON fa.Appointment_Date_Key = ddate.Date_Key
GROUP BY
    ddate.Year,
    fa.Doctor_Surrogate_Id,
    CONCAT(dd.first_name, ' ', dd.last_name);
```

	Year	Doctor_Surrogate_Id	Doctor_Full_Name	Appointments_Count
72	2024	56	Dr. Caleb Owens	13
73	2024	57	Dr. Mira Sharma	1
74	2024	58	Dr. Omar Iqbal	1
75	2025	1	Dr. Alice Miller	10
76	2025	2	Dr. Bob Williams	8
77	2025	5	Dr. Eva Garcia	6
78	2025	11	Dr. Amina Siddiqui	8
79	2025	12	Dr. Javier Morales	2
80	2025	14	Dr. Hiroshi Tanaka	6
81	2025	27	Dr. Anika Reddy	13
82	2025	29	Dr. Clara Oliveira	1
83	2025	30	Dr. Ibrahim Suleiman	1

Query executed successfully.

-- 3. Top Doctors by Appointment Volume

```
SELECT
    CONCAT(d.First_Name, ' ', d.Last_Name) AS Doctor_Name,
    COUNT(fa.Appointment_Id) AS Total_Appointments
FROM
    Fact_Appointments fa
    INNER JOIN Dim_Doctor d ON fa.Doctor_Surrogate_Id = d.Doctor_Surrogate_Id
GROUP BY
    d.Doctor_Surrogate_Id,
    d.First_Name,
    d.Last_Name
ORDER BY
    Total_Appointments DESC;
```

	Doctor_Name	Total_Appointments
1	Dr. Caleb Owens	29
2	Dr. Anika Reddy	24
3	Dr. Alice Miller	22
4	Dr. Sana Ahmed	18
5	Dr. Bob Williams	17
6	Dr. Eva Garcia	15
7	Dr. Amina Siddiqui	15
8	Dr. Hiroshi Tanaka	14
9	Dr. Nia Mensah	12
10	Dr. Maya Torres	7
11	Dr. Mateo Cruz	5
12	Dr. Esme Garcia	5
13	Dr. Ling Kim	2

Query executed successfully.

## Fact\_Billing:

-- 1. Average Bill Amount per Patient in each year.

```
SELECT
    dd.Year,
    dp.Patient_Surrogate_Id,
    CONCAT(dp.First_Name, ' ', dp.Last_Name) AS Patient_Full_Name,
    AVG(fb.Total_Amount) AS Avg_Revenue_Per_Patient
FROM
    Fact_Billing fb
    INNER JOIN Dim_Patient dp ON fb.Patient_Surrogate_Id = dp.Patient_Surrogate_Id
    INNER JOIN Dim_Date dd ON fb.Payment_Date_Key = dd.Date_Key
GROUP BY
    dd.Year,
    dp.Patient_Surrogate_Id,
    CONCAT(dp.First_Name, ' ', dp.Last_Name)
ORDER BY
    dd.Year ASC,
    Avg_Revenue_Per_Patient DESC
```

145 %

Results Messages

	Year	Patient_Surrogate_Id	Patient_Full_Name	Avg_Revenue_Per_Patient
1	2022	36	Mateo Cruz	2500.000000
2	2022	5	Daniel Brown	400.000000
3	2022	8	Sophia Moore	350.000000
4	2022	6	Olivia Martinez	300.000000
5	2022	3	Michael Johnson	300.000000
6	2022	10	Isabella Anderson	300.000000
7	2022	9	David Taylor	280.000000
8	2022	7	James Wilson	250.000000
9	2022	2	Jane Smith	250.000000
10	2022	12	Juan Rodriguez	250.000000
11	2022	33	Maya Torres	250.000000
12	2022	4	Emily Davis	220.000000

✓ Query executed successfully.

-- 2. Total Revenue per Year

```
SELECT
    dd.Year,
    SUM(fb.Total_Amount) AS Total_Revenue
FROM
    Fact_Billing fb
    INNER JOIN Dim_Date dd ON fb.Payment_Date_Key = dd.Date_Key
GROUP BY
    dd.Year
ORDER BY
    dd.Year;
```

212 %

Results Messages

	Year	Total_Revenue
1	2022	7430.00
2	2023	6340.00
3	2024	23680.00
4	2025	16130.00

```
-- 3. Monthly Revenue Trend
```

```
SELECT
    dd.Year,
    dd.Month,
    SUM(fb.Total_Amount) AS Monthly_Revenue
FROM
    Fact_Billing fb
    INNER JOIN Dim_Date dd ON fb.Payment_Date_Key = dd.Date_Key
GROUP BY
    dd.Year, dd.Month
ORDER BY
    dd.Year, dd.Month;
```

193 %

Results Messages

	Year	Month	Monthly_Revenue
1	2022	1	200.00
2	2022	2	400.00
3	2022	3	500.00
4	2022	4	470.00
5	2022	5	580.00
6	2022	6	300.00
7	2022	7	350.00
8	2022	8	2850.00
9	2022	9	480.00
10	2022	10	520.00
11	2022	11	350.00
12	2022	12	430.00

Query executed successfully.

```
-- 4. Top Patients by Total Billing
```

```
SELECT
    dp.Patient_Surrogate_Id,
    CONCAT(dp.First_Name, ' ', dp.Last_Name) AS Patient_Full_Name,
    SUM(fb.Total_Amount) AS Total_Spent
FROM
    Fact_Billing fb
    INNER JOIN Dim_Patient dp ON fb.Patient_Surrogate_Id = dp.Patient_Surrogate_Id
GROUP BY
    dp.Patient_Surrogate_Id,
    CONCAT(dp.First_Name, ' ', dp.Last_Name)
ORDER BY
    Total_Spent DESC;
```

175 %

Results Messages

	Patient_Surrogate_Id	Patient_Full_Name	Total_Spent
1	6	Olivia Martinez	4180.00
2	36	Mateo Cruz	3060.00
3	18	Elijah Okonkwo	3020.00
4	3	Michael Johnson	1350.00
5	5	Daniel Brown	1300.00
6	15	Li Wang	1250.00
7	2	Jane Smith	1250.00
8	9	David Taylor	1090.00
9	12	Juan Rodriguez	1090.00
10	7	James Wilson	1050.00
11	8	Sophia Moore	1050.00
12	29	Zara Khan	1020.00
13	1	John Doe	950.00

Query executed successfully.



## Fact\_Cleaning\_Service:

```
-- 1. Total No. of cleaning services done by each staff member in each year
SELECT
    dd.Year,
    fcs.Staff_Surrogate_Id,
    CONCAT(ds.First_Name, ' ', ds.Last_Name) AS Staff_Name,
    SUM(fcs.Service_Count) AS Total_Cleaning_Services
FROM
    Fact_Cleaning_Service fcs
    INNER JOIN Dim_Date dd ON fcs.Service_Date_Key = dd.Date_Key
    INNER JOIN Dim_Staff ds ON fcs.Staff_Surrogate_Id = ds.Staff_Surrogate_Id
    INNER JOIN Dim_Room dr ON fcs.Room_Surrogate_Id = dr.Room_Surrogate_Id
GROUP BY
    dd.Year,
    fcs.Staff_Surrogate_Id,
    CONCAT(ds.First_Name, ' ', ds.Last_Name)
ORDER BY
    dd.Year,
    fcs.Staff_Surrogate_Id;
```

Results				
	Year	Staff_Surrogate_Id	Staff_Name	Total_Cleaning_Services
1	2022	1	Alice Johnson	5
2	2022	2	Bob Smith	3
3	2022	3	Charlie Brown	4
4	2022	6	Frank Wilson	1
5	2022	13	Michael Taylor	2
6	2022	18	Lina Park	2
7	2022	19	Arjun Singh	2
8	2022	20	Zoe Martinez	2
9	2023	8	Helen Taylor	1
10	2023	13	Michael Taylor	2
11	2023	17	Mateo Garcia	2
12	2024	1	Alice Johnson	6

Query executed successfully.

```
-- 2. Yearly Cleaning Services per Room
SELECT
    dd.Year,
    dr.Room_Number,
    SUM(fcs.Service_Count) AS Total_Services
FROM
    Fact_Cleaning_Service fcs
    INNER JOIN Dim_Room dr ON fcs.Room_Surrogate_Id = dr.Room_Surrogate_Id
    INNER JOIN Dim_Date dd ON fcs.Service_Date_Key = dd.Date_Key
GROUP BY
    dd.Year,
    dr.Room_Surrogate_Id,
    dr.Room_Number
ORDER BY
    dd.Year ASC, Total_Services DESC;
```

Results			
	Year	Room_Number	Total_Services
1	2022	COS302	7
2	2022	EMR802	5
3	2022	PED902	5
4	2022	REH1002	4
5	2023	COS302	4
6	2023	REH1002	1
7	2024	COS302	31
8	2024	REH1002	11
9	2024	PED902	6
10	2024	EMR802	6
11	2024	OPR401	2
12	2024	LAB208	1

Query executed successfully.

-- 3. Total Number of Cleaning Services by Room Type

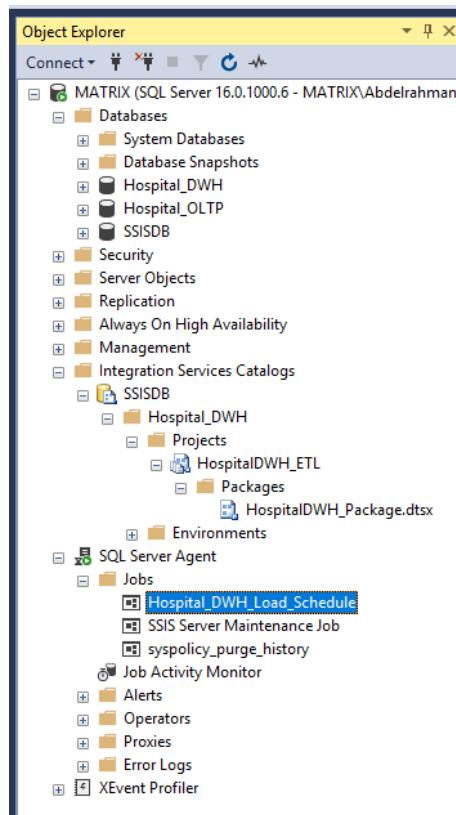
```
SELECT
    dr.Room_Type,
    SUM(fcs.Service_Count) AS Total_Services
FROM
    Fact_Cleaning_Service fcs
    INNER JOIN Dim_Room dr ON fcs.Room_Surrogate_Id = dr.Room_Surrogate_Id
GROUP BY
    dr.Room_Type
ORDER BY
    Total_Services DESC;
```

193 %

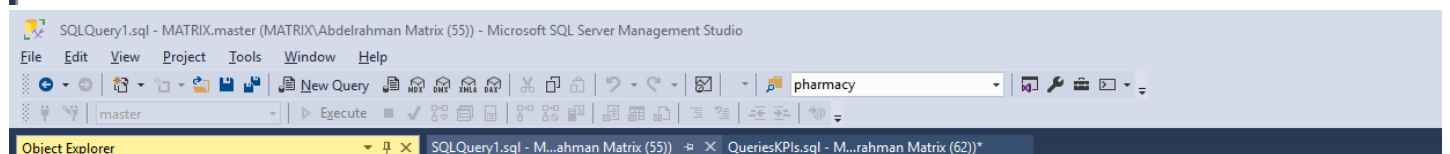
Results Messages

	Room_Type	Total_Services
1	Cosmetic	58
2	Rehabilitation	17
3	Pediatric Ward	12
4	Emergency	12
5	Operating	3
6	General Ward	2
7	Laboratory	2
8	Radiology	2
9	Staff	2

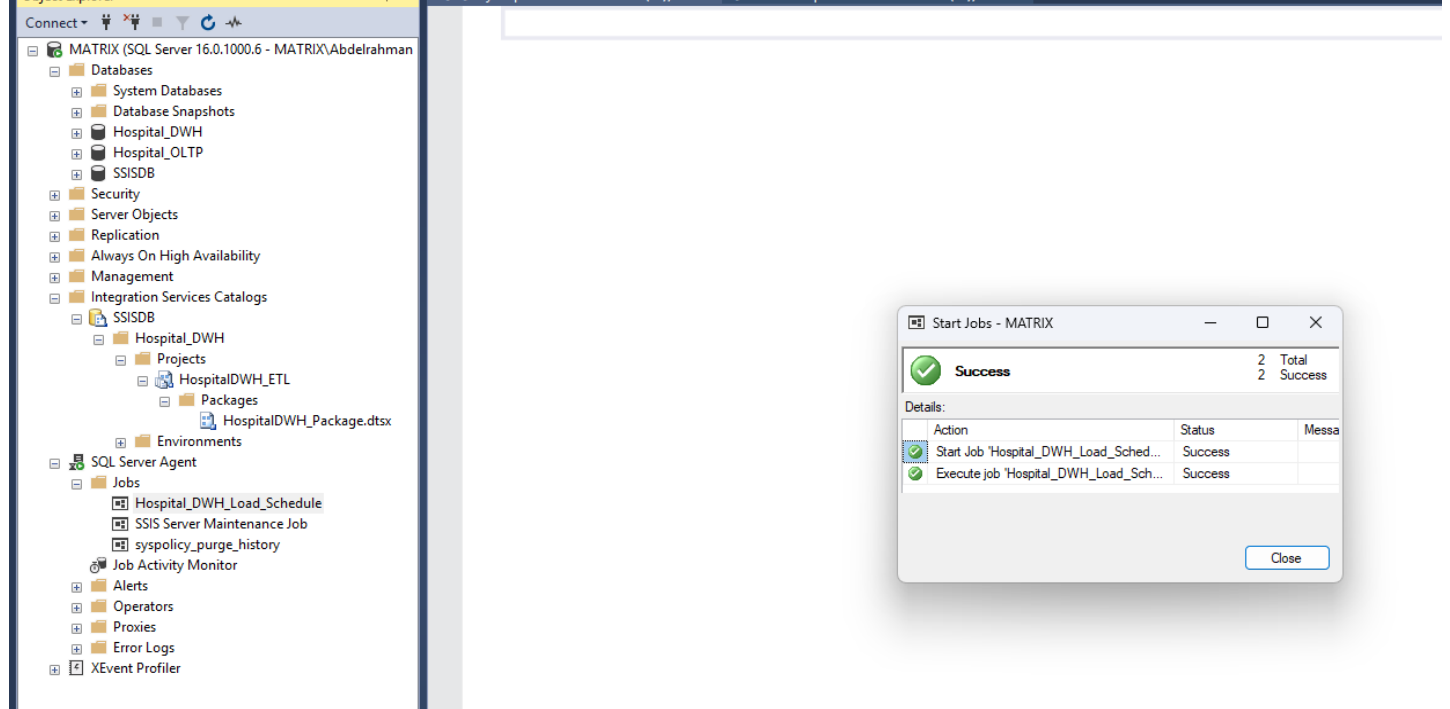
## 6. Screenshots of the deployed packages in SSIS with their schedule:



The screenshot shows the SQL Server Enterprise Manager interface. The left pane displays the Object Explorer tree for the 'MATRIX (SQL Server 16.0.1000.6 - MATRIX\Abdelrahman)' server. The 'SSISDB' folder is expanded, showing the 'Hospital\_DWH' folder, which contains 'Projects' and 'Packages'. The 'Hospital\_DWH\_Load\_Schedule' job is highlighted under the 'Jobs' folder.



The screenshot shows the SQL Query window in Microsoft SQL Server Management Studio. The query window is titled 'SQLQuery1.sql - MATRIX.master (MATRIX\Abdelrahman Matrix (55)) - Microsoft SQL Server Management Studio'. The query text is 'pharmacy'. The Object Explorer on the left shows the 'MATRIX (SQL Server 16.0.1000.6 - MATRIX\Abdelrahman)' server, with the 'Hospital\_DWH\_Load\_Schedule' job highlighted.

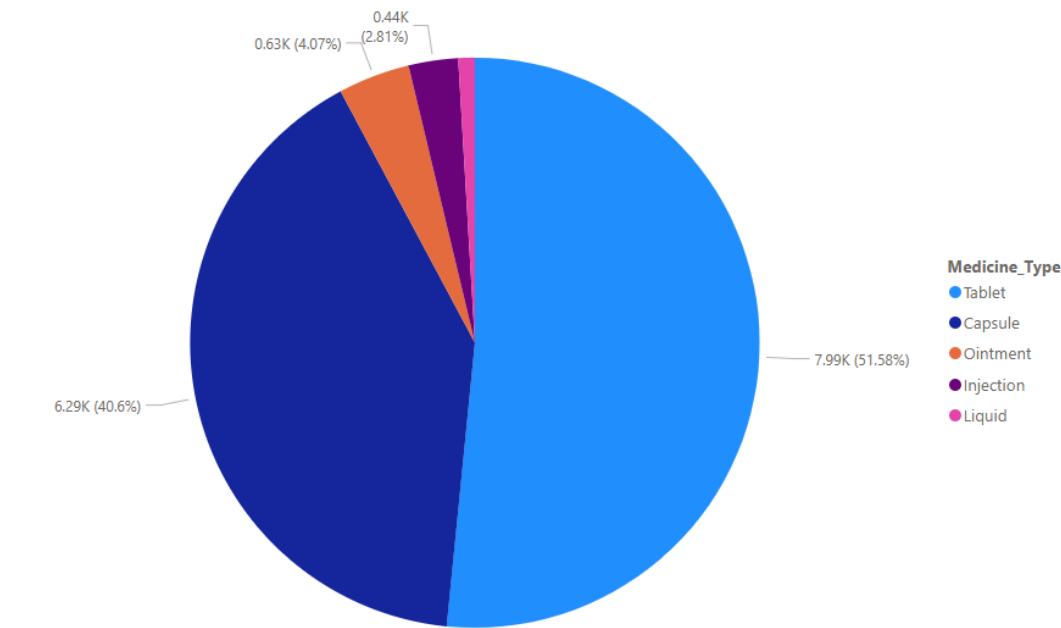


The screenshot shows the 'Start Jobs - MATRIX' dialog box. The dialog box displays the status of the 'Hospital\_DWH\_Load\_Schedule' job. The status is 'Success'. The details table shows the following information:

Action	Status	Message
Start Job 'Hospital_DWH_Load_Sched...	Success	
Execute job 'Hospital_DWH_Load_Sch...	Success	

7. [Bonus] Build an interactive dashboard for the DWH using any data visualization tool (Ex: Microsoft Power BI).

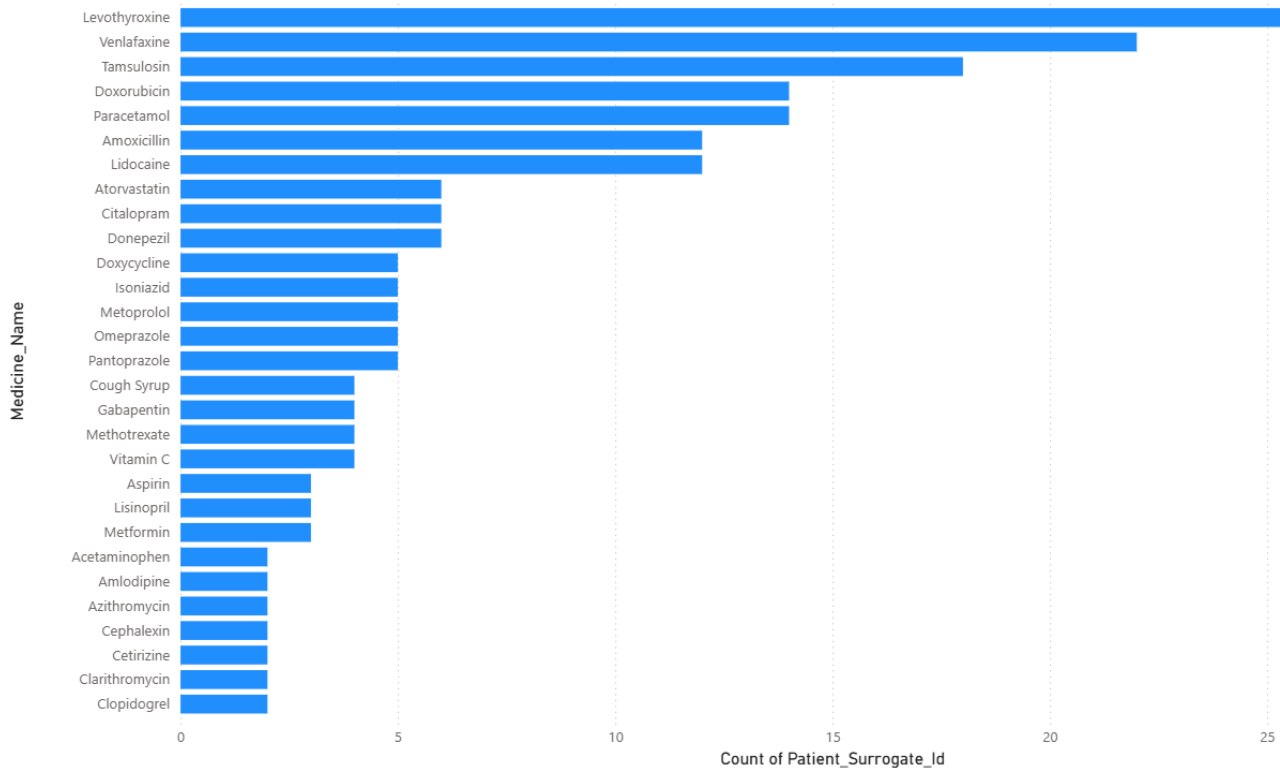
Sum of Total\_Quantity\_Dispensed by Medicine\_Type



Dim_Medicine.Medicine_Name	2022	2023	2024	2025	Total
Acetaminophen				45.00	45.00
Amlodipine	110.00				110.00
Amoxicillin		20.00	220.00	160.00	400.00
Aspirin		30.00	60.00		90.00
Atorvastatin	100.00	50.00	165.00	100.00	415.00
Azithromycin		160.00			160.00
Cephalexin	35.00				35.00
Cetirizine			90.00		90.00
Ciprofloxacin			25.00		25.00
Citalopram	20.00	25.00	40.00	20.00	105.00
Clarithromycin				70.00	70.00
Clopidogrel	20.00	60.00			80.00
Cough Syrup	5.00	5.00	10.00		20.00
Dexamethasone				50.00	50.00
Donepezil	270.00		360.00		630.00
Doxorubicin		10.00	170.00	80.00	260.00
Doxycycline	50.00	50.00	150.00	150.00	400.00
Duloxetine	80.00				80.00
Epinephrine			50.00		50.00
Escitalopram				50.00	50.00
Esomeprazole	120.00				120.00
Ferrous Sulfate			65.00		65.00
Fluoxetine		45.00			45.00
Furosemide	100.00				100.00
Gabapentin		120.00	50.00		170.00
Hydrochlorothiazide		50.00	30.00		80.00
Hydrocortisone			90.00		90.00
Ibuprofen			15.00		15.00
Insulin Glargine		60.00	15.00		75.00
Isoniazid	50.00	50.00	120.00	50.00	270.00
Lamotrigine			15.00		15.00
Levofloxacin	120.00				120.00
Total	2,175.00	1,415.00	6,845.00	5,045.00	15,480.00

< Back to report

COUNT OF PATIENT\_SURROGATE\_ID BY MEDICINE\_NAME

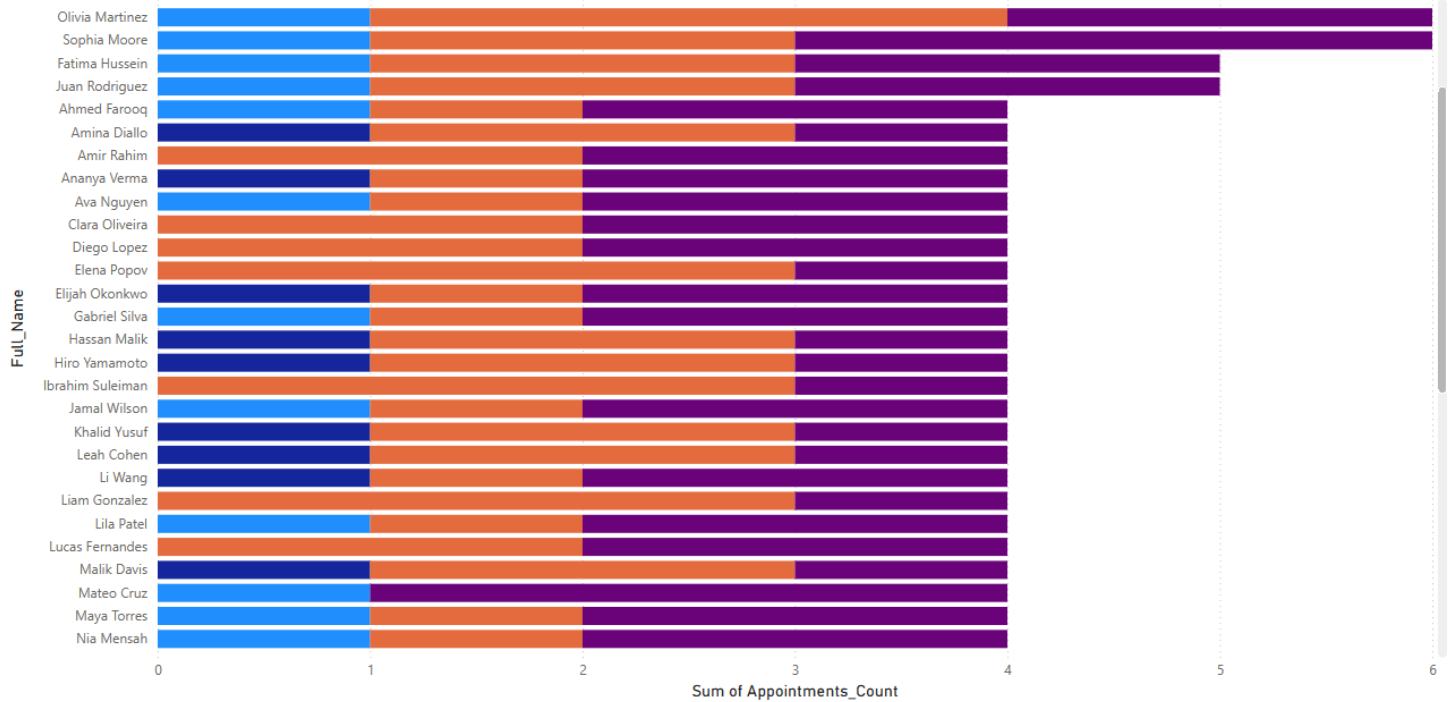


↑ ↓ ⇅ ↺ ≡ ⋮

< Back to report

SUM OF APPOINTMENTS\_COUNT BY FULL\_NAME AND YEAR

Year ● 2022 ● 2023 ● 2024 ● 2025



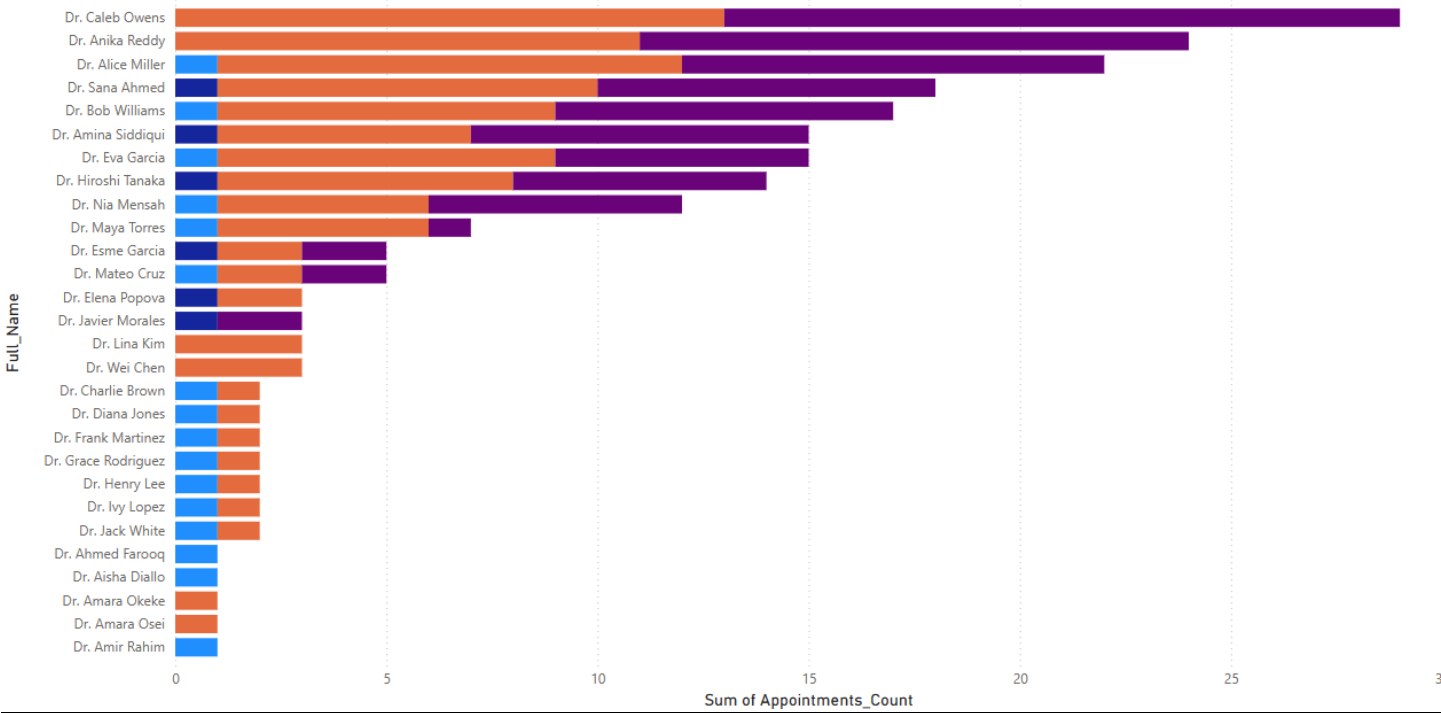


< Back to report

SUM OF APPOINTMENTS\_COUNT BY FULL\_NAME AND YEAR

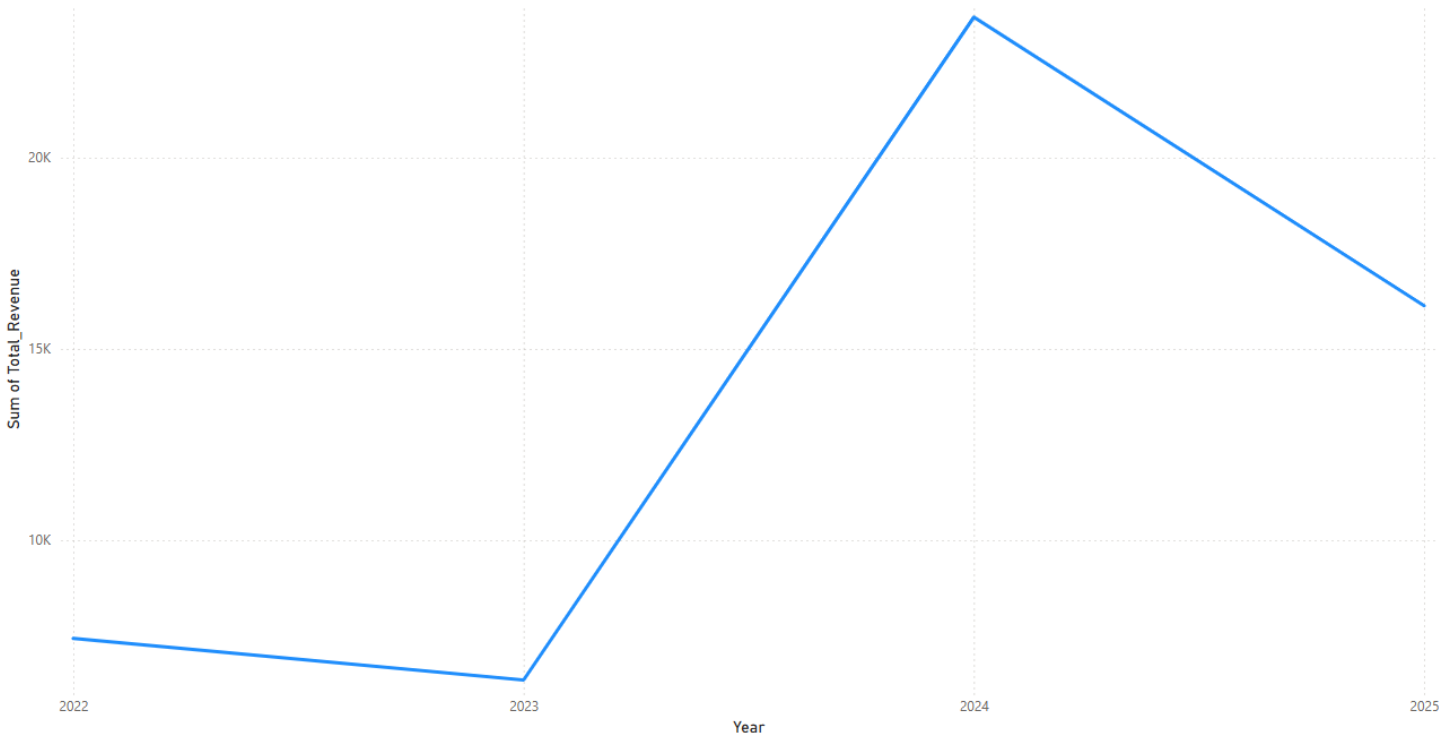


Year 2022 2023 2024 2025

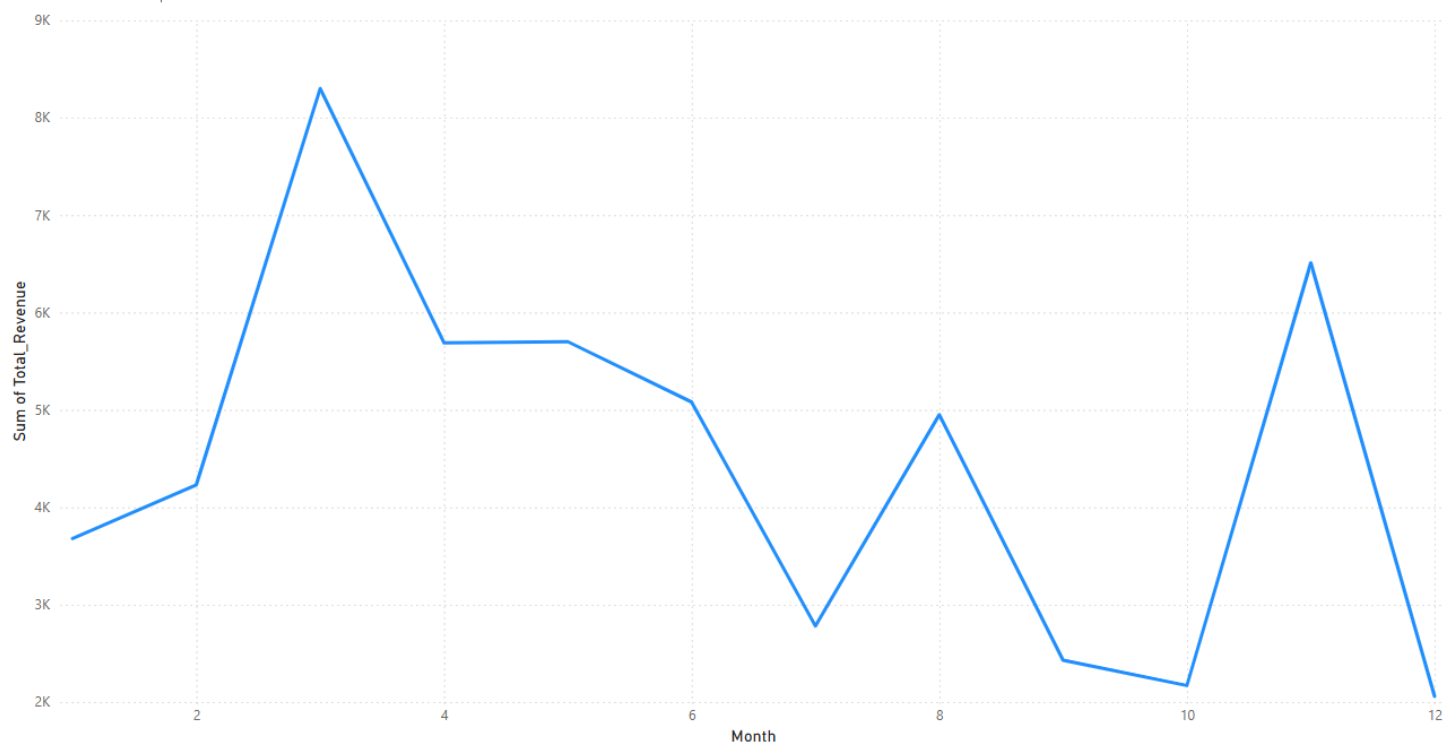


< Back to report

SUM OF TOTAL\_REVENUE BY YEAR



< Back to report | SUM OF TOTAL\_REVENUE BY MONTH



< Back to report | SUM OF TOTAL\_SERVICES BY ROOM\_TYPE

